

开发专家
之 数据库



朱德利

飞思科技产品研发中心

编著
监制

SQL Server 2005 数据挖掘与商业智能 完全解决方案

轻举技术之“纲”，张合应用之“目”，锻造SQL Server 2005最全面的商业智能指南
从理论到实践，凝聚微软主流数据库最前沿的技术要领，本书将深入浅出讨论：

商业智能、数据仓库、数据挖掘、部署、管理、SSIS、SSAS、SSRS、OLAP、
KPI分析、MDX、数据库、80/20法则



电子工业出版社

PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

包含书中实例源文件



开发专家
之 数据库

朱德利
飞思科技产品研发中心

编著
监制

SQL Server 2005 数据挖掘与商业智能 完全解决方案

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内容简介

本书以 BI 解决方案的体系结构为中心,以 SQL Server 2005 为载体,将着眼点放在数据挖掘和商业智能上,详细讲解了数据报表、数据分析和数据挖掘这 3 个层面在 SQL Server 中的理论和技术细节。

全书共 12 章,全面阐述了 SQL Server 2005 商业智能平台中 SSIS、SSAS 和 SSRS 的使用技巧和在商业智能领域的应用方法。本书内容分为 3 个层次。前两章为第 1 层次,是基础,第 3 章开始到第 10 章为第 2 层次,是 OLAP 应用,最后两章为第 3 层次,是数据挖掘。其中,第 1 章是对 BI、DW、OLAP 和 DM 的基本概念和理论的综述。第 2 章创建了一个完整的 BI 应用示例。第 3 章讲述了设计一个结构良好的数据仓库的方法和技巧。第 4 章和第 5 章则详述了数据整合工具 SSIS 的使用方法及其在 BI 领域的应用范例。第 6 章和第 7 章针对分析服务 SSAS 的基本使用和高级使用进行了讲解。第 8 章描述了 MDX 在多维数据库中的应用。第 9 章描述的是用 SSRS 处理智能报表的技术。第 10 章介绍了前面的知识在商业智能分析中的综合应用。第 11 章和第 12 章描述的是数据挖掘技术在 SQL Server 平台下的基本使用和满足商务分析需求的具体示例。

本书内容翔实,示例丰富,结构合理,语言简洁流畅。在写作过程中力求把每一个知识点和技术方法讲深讲透。本书良好的结构设计保证了它既可以作为各种数据库培训班和大专院校的数据库、数据仓库和数据挖掘领域的教材,又可作为各类开发人员及企业管理人员的参考用书。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

SQL Server 2005 数据挖掘与商业智能完全解决方案 / 朱德利编著. —北京: 电子工业出版社, 2007.10
(开发专家之数据库)

ISBN 978-7-121-05015-2

I. S… II. 朱… III. 关系数据库—数据库管理系统, SQL Server 2005 IV. TP311.138

中国版本图书馆 CIP 数据核字 (2007) 第 140731 号

责任编辑: 王树伟

印 刷: 北京智力达印刷有限公司

装 订: 北京中新伟业印刷有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本: 787×1092 1/16 印张: 25.25 字数: 646.4 千字

印 次: 2007 年 10 月第 1 次印刷

印 数: 5 000 册 定 价: 46.00 元 (含光盘 1 张)

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线: (010) 88258888。

从技术到工程，软件工程师需要顺利从编码者成长为系统设计者，也需要在进行系统设计的同时掌握软件需求、测试等环节的方法。

出版说明

近两年系统开发和应用的一些趋势

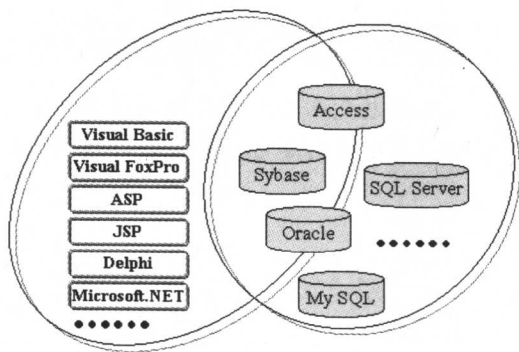
近几年，在数据库系统开发或系统集成中，有几种态势的发展越来越清晰，值得我们关注，其中也包含技术和应用方面的一些变化：

- 软件工程方法越来越受到重视，并且从烦琐的“国际模板”中逐步解脱出来，总结出了一套适合我们自己的软件工程的“简化版本”，没有令人眼花缭乱的概念，也没有没完没了的控制和反馈，也没有无穷无尽的单据需要确认。
- IT应用格局从C/S体系的一支独秀，发展到C/S和B/S共分天下。这可能是近年来IT应用方面最大的一个变化。这两种类型的系统其实在满足特定的客户需求上毫无优劣高下之分，适合的就是最好的。系统设计师不应该也不可能因为自己的喜好，而不顾客户的需求来设计软件系统。
- 涌现了很多新的技术和新型的系统应用。事实上，只有新的需求才能刺激新的应用诞生，绝对不会因为新的技术而带来新的应用。清醒的程序员把精力放在系统需求的变化上，盲动的程序员把精力放在不停地追赶技术的更迭上。

关于《开发专家之数据库》丛书

《开发专家之数据库》丛书是在充分把握技术人员需求脉搏的基础上，强力打造的新作品。本丛书在延续原有优势的基础上，针对行业系统开发特征和数据库开发的不同技术，融合了当前主流的程序开发语言和主流的数据库系统，为不同行业、不同工具掌握者提供最实用的开发指南。

数据库开发一直是软件编程中最受关注的领域。长久以来，各种在开发领域中雄踞一时的语言或技术，都和数据库开发有非常密切的关系。软件研发人员在选择开发语言和数据库时，就像在做一個配对游戏，每个软件公司和每个开发者都在不断地改变着他们的选择。在他们不断进行语言和工具的选择的同时，社会在变化，行业的需求在不断地扩展。可以说，行业信息化的热潮才刚刚开始，特别是中、小型企业的信息化建设步伐刚刚迈出第一步，这个软件开发的市场是非常巨大的，迫切需要大量高素质的开发从业人员。



行业应用系统的开发有着它独有的特征和要求。一个成熟的软件开发者，必然是深知行业需求、胜任于多种构架的系统设计、娴熟于各类开发工具的多面手。从根本上说，这种三位一体的完美结合，才能创造出一个合格的行业系统。当然，我们也不能忘记周边还有许多工作要做，比如系统测试等。

如何突破：新案例、新应用、新方法

第一批“开发专家之数据库”丛书出版以来，得到了广大读者的认可。我们感到有必要在信息系统开发领域推出新的图书，来帮助另外一些有了一定技术基础的读者顺利过渡到真实的软件工程思维中来，这就是我们所定义的“突破”的含义。

归纳起来，第二批图书也将从行业角度出发，整合各类开发语言和数据库系统，把系统设计、语言编程、软件测试等规范的软件工程无缝嵌入到本丛书的叙述中，为读者提供一整套行业系统开发的解决方案。

	C/S 体系	B/S 体系	综合与融合
I. 第一层面 (第一批图书) → 技术基础 + 案例，完成从入门到提高的过渡，把握软件工程方法	《Visual Basic 数据库开发实例精粹》(第 2 版)	《JSP 数据库开发实例精粹》	《SQL Server 数据库开发实例精粹》
	《Visual FoxPro 数据库开发实例精粹》	《ASP 数据库开发实例精粹》(第 2 版)	《SQL Server 2005 数据库管理与应用高手修炼指南》
	《Visual Basic.NET 数据库开发实例精粹》	《ASP.NET 2.0 数据库开发实例精粹》	《SQL Server 2005 数据挖掘与商业智能完全解决方案》
II.第二层面 (第二批图书) → 纯案例，深入软件工程方法，并提供周边工具和系统配置的多媒体教程	《信息系统开发实例精粹 (Visual Basic 版)》	《网络系统开发实例精粹 (JSP 版)》	
	《信息系统开发实例精粹 (C#版)》	《网络系统开发实例精粹 (ASP 版)》	
		《网络系统开发实例精粹 (ASP.NET 版)》	
不断修订、再版，以更新技术，更新案例，优化内容。			

本丛书大部分图书特别为读者开发了多媒体教学光盘，旨在提供给大家更快捷的学习途径、更丰富的学习内容，能够通过一张光盘学到更全面的技术。

电子工业出版社计算机研发部

联系方式

咨询电话: (010) 68134545 88254160

电子邮件: support@fecit.com.cn

服务网址: <http://www.fecit.com.cn> <http://www.fecit.net>

通用网址: 计算机图书、飞思、飞思教育、飞思科技、FECIT

An organization's ability to learn, and translate that learning into action rapidly, is the ultimate competitive advantage.

Jack Welch

Chairman, General Electric

这是被誉为“世界第一 CEO”的杰克·韦尔奇的一句话：一个组织的学习能力，以及快速地把所学到的东西转化为行动的能力，才是这个组织的终极竞争优势。

杰克·韦尔奇在 20 年内把一个机构臃肿，等级森严，对市场反应迟钝，并且在全球竞争中正走下坡路的 GE 变成“全美最受推崇公司”。笔者经常思考这句话，对于一个人来说，只要定位清楚、有理想、有远见，打造出有自己竞争优势的学习内容和学习方法应该不是一件难事，但是对于一个组织来说，这个“终极竞争优势”应该如何打造，却不是件容易的事。

“物竞天择，适者生存”是大自然的一个普遍法则，不仅适用于生物体，也适用于企业。而今置身于市场经济且面向全球性激烈竞争的环境下，任何商家的优势都不是单纯地取决于产品、服务和地区等方面的因素，而在于知识，在于创新。谁最先获得有用的知识并转化为行动，谁就可能成为赢家。随着企业信息化的普及，承载信息的数据随着时间的推移而不断积累并与日俱增，使得企业难以从大量数据中找到真正有用的决策信息。因此，企业迫切需要新的技术和方法从大量数据中提取有价值的信息和知识。数据仓库、数据挖掘及建立在它们基础之上的商业智能就是这种技术之一。数据仓库是一种针对大量数据集中进行数据组织与管理的技术，专门用于支持分析型的数据查询，而数据挖掘是从大量数据中寻找蕴涵在其中但尚未被发现的知识，而商业智能则是把这两者应用于商务领域，提高决策效率。

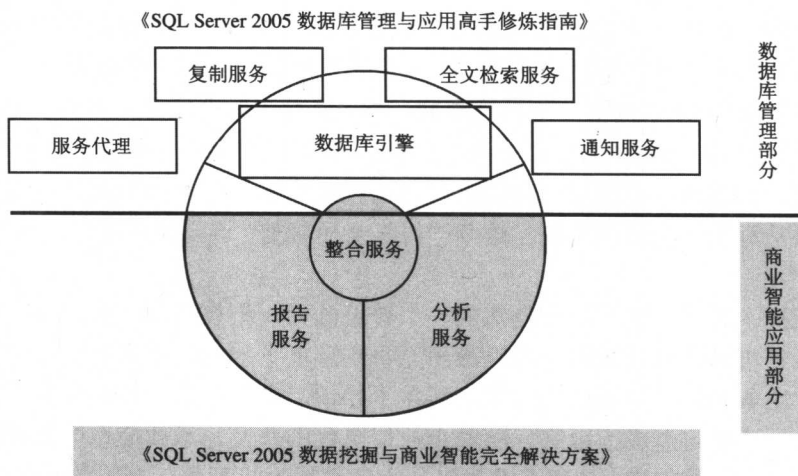
在全球激烈的市场竞争中，只有在市场上领先的企业，才能立于不败之地。商业智能和数据挖掘越来越受到企业的重视，要打造组织的“终极竞争优势”，这应该是一条必经之路。

“社会一旦有技术上的需要，则这种需要就会比十所大学更能把科学推向前进。”正是因为对商业智能和数据挖掘的需求已经成为了一种普遍共识，各大软件厂商才对此趋之若鹜。微软也不例外，在其推出的新版本数据库管理软件 SQL Server 2005 中，不仅在数据管理方法和数据库应用程序开发方面有了很大的改进，在商业智能方面更是发生了巨大的变化。笔者在 BI 和 DM 领域已经工作多年，在 SQL Server 2005 推出后，第一时间使用了其增强的 BI 功能，并在学习和应用中有所感悟。笔者认为有必要把这些感悟组织成文，提供给同行和初学者，于是有了本书。以下是关于本书在这些方面的一些说明。

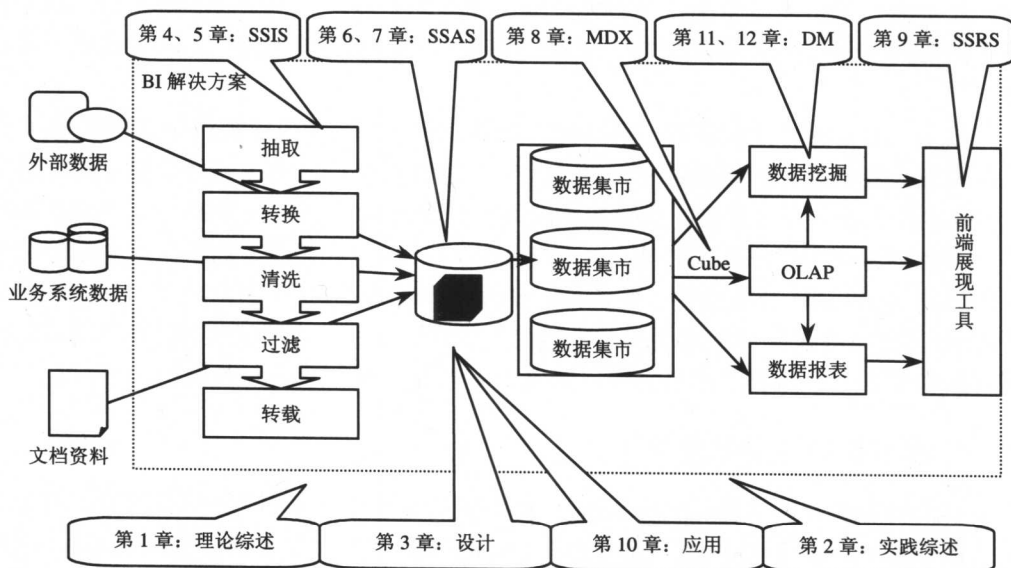
一、本书的内容安排思路

本书的内容和其姊妹篇《SQL Server 2005 数据库管理与应用高手修炼指南》相比，实际上是把 SQL Server 2005 的应用从数据的技术实现提升到了数据的价值实现。除了数据管理方面的组件外，在商业智能方面，SQL Server 2005 还提供了分析服务(Analysis Services)、

报表服务 (Reporting Services) 和数据挖掘 (Data Mining) 的全面集成, 数据转换服务 (DTS) 由 SQL 服务器集成服务 (SSIS) 代替。基于 SQL Server 2005 的这些内部构成原理, 本书将着眼点放在数据挖掘和商业智能上, 数据库的管理与应用则作为阅读本书的基础。具体的内容安排如下图所示, 阴影部分即为 SQL Server 2005 商业智能和数据挖掘部分, 也是本书重点论述的对象。



数据报表、数据分析和数据挖掘是 BI 的 3 个层面。不管是用什么软件作为商业智能的平台, 一个完整的 BI 应用都遵循统一的一个流程, 这就是 BI 解决方案的体系结构, 如下图所示。本书在安排基于 SQL Server 的 BI 系统开发技术内容时, 把这个体系结构作为本书结构构成的依据, 目的就是希望读者随时都能以系统的眼光来看待 BI 构建中的每一个细微环节。



从图中可以看到，全书共 12 章，每一章都可以在 BI 解决方案体系结构中找到相应的坐标。具体而言每一章的内容是：

第 1 章 发掘数据金矿的工具：BI 与 DW、OLAP、DM。先简单地说明企业经营管理活动对商业智能的需求，从而明确实施 BI 的原因，解决 Why 的问题；再阐述商业智能的技术构成，大概了解如何实现商业智能，解决 What 和 How 的问题；最后列举一些实际的部署商业智能项目结果，解决 What Effect 的问题。

第 2 章 构建简单的 BI 应用：福马特商业智能系统。在内容安排上属于实践中的“务实”，理论上的“务虚”，只着眼于操作过程而忽略其中的深奥理论，目的在于明确一个商业智能应用的全貌。

第 3 章 BI 分析的基石：结构良好的数据仓库设计。对数据仓库的设计进行深入学习，把复杂的数据仓库与商业智能理论融入到具体的操作中，探究构建数据仓库的理论和方法。

第 4 章 用 SSIS 对数据进行 ETL 操作。学习针对数据仓库的提取、转换和加载操作。包括 SSIS 关键元素的使用、一个完整的 SSIS 包的创建方法、包的调试、包的配置和部署以及 SSIS 的管理。

第 5 章 SSIS 在商业智能中的典型应用。如何用 SSIS 工具来满足商业智能系统中复杂的数据整合需求。包括 SSIS 在 BI 系统中的应用场景，一个复杂的商务应用实例分析，业务数据装载，数据清洗，另外还对 SSIS 商业智能的应用进行了总结。

第 6 章 用 SSAS 进行 OLAP 操作。这一章的任务就是去实践 OLAP 理论，并进一步把它转化为实际应用。数据仓库最重要的一个组成部分就是数据立方，因此，数据立方的设计技巧和操作方法也理所当然地成为本章的主要内容。

第 7 章 数据立方的增强及其应用。包括对维度与度量关系的更改、为 Cube 增加计算来增强 Cube 的功能、定义和使用 KPI 来满足商业需求等。

第 8 章 用 MDX 扩展 OLAP 功能。MDX 在多维数据库中的应用，不仅为我们提供了便利查询和获取多维数据的有效途径，还扩展了多维数据的视觉化效果，从而使我们的商业智能系统更加符合企业应用的需求，实现强大的功能。本章将针对能够对多数数据进行访问和操作的 MDX 语言，从简单的语法要素入手，以进阶的方式，学习 MDX 从基础到高级的应用。

第 9 章 用 SSRS 处理智能报表。本章是前面商业智能应用解决方案的继续，把报表服务作为 SQL Server 2005 提供的 BI 平台的前端展现部分进行描述。包括报表的基本知识、SSRS 的配置与管理、基本报表的增强和报表生成器的使用等内容。

第 10 章 基于 SSAS 的商业智能分析。本章的目的是在前面讲述的技术和管理需求之间搭建一座桥梁，使得技术真正能够用来解决商务活动中的实际问题。包括构架商业智能系统的生命周期、商业智能向导、KPI 分析、80/20 法则的分析和用专业前端展现工具呈现商业信息。

第 11 章 数据挖掘体系结构与基本使用方法。进入商业智能更深层次的应用——数据挖掘。包括数据挖掘的基本知识、SQL Server 数据挖掘方案的构成和一个完整的数据挖掘示例——线性回归算法（逻辑回归算法，决策树算法，聚类分析算法，Naive Bayes 算法

和关联算法的使用), 还对数据挖掘进行了简单地总结。

第 12 章 用数据挖掘技术满足商业分析需求。包括商业智能应用中的数据挖掘、潜在客户分析、购物篮分析和数据挖掘的前端展现等知识。

二、本书的写作特色

本书定位在实践上面, 而不是从理论到理论。但是作为商业智能必须应用和管理活动中, 管理理论在 BI 项目的成功实施里占据了十分重要的地位。基于此, 本书遵循一种“理论—实践—理论”的学习路径, 即先概要认识理论的实践价值, 再在使用技术实现此理论的过程中进行体会。沿着这条路径, 我们为本书具有以下几方面的特点做了坚持不懈的努力。

1. 实践的厚度

商业智能不是夸夸其谈, 而是产生效益、数据挖掘不是数学公式, 而是找到模式。本书着眼于实践, 具体讲述了业务数据装载、数据清洗、关键性能指标分析方法和 8/2 法则的分析方法等, 真正能满足商务需求的操作, 并力求对实现过程和操作步骤的表述清晰、简洁。在数据挖掘方面, 本书避免了复杂的算法实现和推理, 而是在简明扼要的论述其算法精髓的基础上, 重点讲解如何通过技术手段实现这个算法在商务领域的应用, 同时还讲述了潜在客户分析和购物篮分析等具体技术细节。

2. 理论的深度

理论可以让实践产生更好的效果。BI 的特殊性决定了经济管理理论在技术实践过程中的巨大指导作用。因此本书以 BI 解决方案的体系结构为全书的骨架, 具体描述了每一个技术实现细节的商务需求和管理要点。可以说, 这在所有 BI 平台中都是通用的。通过本书的学习, 可以系统地领悟商业智能和数据挖掘在解决实际问题中的各个方面。

3. 学习的效度

笔者在数据库及其应用领域研究多年, 深感把数据的价值应用于商务中相关知识的不系统性, 因此在写作过程中很重视知识体系的完整性和学习曲线的平缓性, 体现在行文中, 本书每一章都有一个“本章导读”来导航该章的学习, 同时在知识跨度和难度比较大的章节都有总结, 这些总结都是笔者在学习和工作中心血的结晶, 可以为读者有效地学习提供很大的帮助。

基于以上的这些努力, 我们希望借微软推出的 SQL Server 2005 这个产品, 把本书写成 BI 和 DM 领域有理论、有实践、既通俗、又经典的书。能不能达成所愿, 还需要时间和读者的检验。

本书是我的呕心沥血之作, 在成书的过程中我牺牲了很多东西, 历时也较长。我不希望它纯粹是一本畅销书, 而是希望它能历久弥香, 以此为标准而成此书, 着实费了一番心血。

在写作过程中, 得到了许多人的帮助, 特别是爱妻徐铭莲女士, 没有你的全力支持, 我是不可能有任何作为的, 辛苦你啦!

本书参考了许多先贤圣达的资料和文章，大部分都在参考文献中列出来了，借此机会对他们表示衷心的感谢。

另外，还要特别感谢的是电子工业出版社的田小康先生，在成书过程中，你以大海般的胸襟原谅了我许多过失，没有你的这份理解和宽容，也就没有本书顺利的写作。

尽管在写作本书的过程中，作者和编辑都尽可能地字斟句酌，尽量减少错和漏，但毕竟人无完人，金无足赤，本书也不可避免地会出现一些 bug，欢迎广大读者批评指正！同时针对此书的内容和技术上的任何问题，作者将会提供不间断地持续增值服务，读者要做的只是把相关内容发送到 zdlxml@126.com 即可。

重庆师范大学 朱德利

2007 年 9 月

目 录

第1章 发掘数据金矿的工具:

BI 与 DW、OLAP、DM.....	1
1.1 企业经营管理活动对 商业智能的需求	1
1.1.1 企业经营活动面临的挑战	1
1.1.2 企业决策实现过程的 信息需求	4
1.1.3 企业信息化系统的进化	5
1.2 商业智能的技术构成	7
1.2.1 什么是商业智能	7
1.2.2 商业智能的结构描述	9
1.2.3 数据挖掘和商业智能工具	11
1.2.4 商业智能工具的选择	12
1.2.5 SQL Server 2005 的 商业智能构架	13
1.3 部署商业智能	14
1.3.1 商业智能如何协助 企业管理	14
1.3.2 商业智能在各领域的应用	15
1.3.3 商业智能应用实例	17

第2章 构建简单的 BI 应用:

福马特商业智能系统	21
2.1 设计和创建数据仓库	21
2.1.1 原始业务数据分析	21
2.1.2 设计数据仓库逻辑模型	22
2.1.3 创建 foodmartsaleDW 数据仓库	23
2.2 设计和使用 ETL	23
2.3 创建 OLAP 数据立方	27
2.3.1 定义数据源	27
2.3.2 定义数据源视图	28
2.3.3 生成多维数据集	30
2.4 创建和使用报表	34
2.4.1 创建报表	35
2.4.2 使用报表	38

2.5 实现其他前端展现	39
2.6 使用数据挖掘获取商业智能	42
2.6.1 商务需求分析	42
2.6.2 创建挖掘结构	43
2.6.3 从数据挖掘中获取 有价值的信息	44
第3章 BI 分析的基石: 结构良好的数据仓库设计	47
3.1 数据的两种组织形式: 操作数据和分析数据	47
3.1.1 操作型系统和分析型 系统的分离	47
3.1.2 事务处理和分析 处理的对比	48
3.1.3 操作型数据与分析型 数据的对比	49
3.1.4 数据仓库的特点	50
3.2 数据仓库设计方法论	53
3.2.1 数据库设计与 数据仓库设计	53
3.2.2 数据仓库的架构方式 及其比较	55
3.2.3 宏观上的数据仓库设计	60
3.2.4 微观上的数据仓库设计	61
3.2.5 2 种创建数据仓库的模式	62
3.2.6 技术上需要关注的 重点步骤	63
3.3 理解历史数据和分析需求	64
3.3.1 “数据驱动 + 用户驱动” 的设计理念	64
3.3.2 理解业务数据	64
3.3.3 确定用户对分析型 数据的需求	69
3.4 明确仓库的对象: 主题和元数据	72
3.4.1 信息打包技术	73

Contents

3.4.2 理解数据仓库中的主题	77	4.2.4 优先约束	136
3.4.3 理解数据仓库中的元数据	80	4.2.5 源	138
3.5 确定分析内容的构成:		4.2.6 转换	139
事实及其粒度	81	4.2.7 目标	140
3.5.1 事实、度量和事实表	82	4.2.8 连接管理器	141
3.5.2 事实表的设计	82	4.2.9 变量	141
3.5.3 粒度的设计	83	4.2.10 事件处理程序	142
3.5.4 聚合的设计	89	4.2.11 日志提供程序	143
3.5.5 数据分割	90	4.3 创建一个完整的 SSIS 包	146
3.6 规划分析的视角: 维度	91	4.3.1 生成 SSIS 解决方案	147
3.6.1 维度的构成	91	4.3.2 设计控制流	148
3.6.2 维度的特性	91	4.3.3 设计数据连接	150
3.6.3 维度的分类	92	4.3.4 设计数据流	151
3.6.4 维度的层次和级别	94	4.3.5 项目总结	155
3.6.5 维度的缓慢变化 特性及其处理	95	4.4 包的调试	155
3.6.6 典型的维度设计	97	4.4.1 控制流的调试	156
3.7 数据仓库物理模型设计	99	4.4.2 数据流的调试	157
3.7.1 设计存储结构	99	4.5 包的配置和部署	159
3.7.2 设计索引策略	100	4.5.1 包的配置	159
3.7.3 设计存储策略	100	4.5.2 包的部署	161
3.8 数据仓库设计示例	102	4.5.3 包的运行	162
3.8.1 销售数据仓库	102	4.6 SSIS 的管理	164
3.8.2 保险业数据仓库	103	4.6.1 管理 SSIS 服务	164
3.9 数据仓库数据库设计的心得总结	104	4.6.2 配置 SSIS 服务	165
3.9.1 透彻理解数据 仓库设计过程	104	第 5 章 SSIS 在商业智能中的典型应用	167
3.9.2 把握设计的关键环节	104	5.1 SSIS 在 BI 系统中的应用场景	167
3.9.3 分离非分析数据	105	5.1.1 合并异类数据	167
第 4 章 用 SSIS 对数据进行 ETL 操作	107	5.1.2 填充数据仓库和数据集市	167
4.1 认识 SSIS	107	5.1.3 数据清洗	168
4.1.1 使用 SSIS 的效果	107	5.1.4 处理过程中加入智能转换	168
4.1.2 SSIS 的工作原理	108	5.1.5 自动化数据管理和加载	168
4.1.3 第一个 SSIS 包的设计	110	5.2 一个复杂的商务应用实例分析	169
4.2 SSIS 关键元素的使用	118	5.2.1 包的构成	169
4.2.1 包	118	5.2.2 功能及其实现的原理分析	170
4.2.2 容器	119	5.3 业务数据装载	175
4.2.3 任务	126	5.3.1 商务需求及其设计思路	175
		5.3.2 设计基本的功能	177
		5.3.3 用循环加载全部数据	179

5.3.4 增强数据加载的灵活性	179	7.1.2 修改“客户”维度	232
5.3.5 增强包的可维护性	181	7.1.3 修改“时间”维度	234
5.3.6 增强包的鲁棒性	182	7.1.4 修改“产品”维度	236
5.4 数据清洗	183	7.2 加入分销事实表及其维度	238
5.4.1 商务需求	184	7.3 完善新增维度	239
5.4.2 设计思路	185	7.3.1 父子维度的使用	239
5.4.3 方案实现	185	7.3.2 分销商维度的修改	242
5.5 SSIS 商业智能应用总结	187	7.3.3 雇员维度的修改	243
第6章 用 SSAS 进行 OLAP 操作	189	7.3.4 促销维度的修改	244
6.1 AS2005 中 OLAP 的设计方法	189	7.3.5 维度属性的特殊处理	244
6.1.1 自下而上的设计方法	189	7.4 设置维度与度量值组的关系	248
6.1.2 自上而下的设计方法	190	7.4.1 常规维度关系及其定义	249
6.1.3 两种方法的使用	191	7.4.2 引用维度关系及其定义	250
6.2 统一维度模型	191	7.4.3 事实维度关系及其定义	252
6.3 OLAP 的操作对象:		7.4.4 多对多维度关系及其定义	254
数据立方的创建	193	7.5 增加计算	257
6.3.1 自上而下生成 Cube	194	7.5.1 计算成员的创建	257
6.3.2 自下而上生成 Cube	203	7.5.2 命名集的创建	260
6.3.3 部署 Cube	207	7.5.3 其他脚本命令的创建	261
6.4 Cube 的构成及其各种操作	208	7.6 定义和使用 KPI	261
6.4.1 Cube 的构成	209	7.6.1 KPI 的设计	262
6.4.2 多维数据集结构	210	7.6.2 KPI 的使用	265
6.4.3 维度用法	213	7.7 增强操作	265
6.4.4 计算	214	7.8 配置访问权限	267
6.4.5 KPI	215	第8章 用 MDX 扩展 OLAP 功能	269
6.4.6 操作	216	8.1 初识 MDX	269
6.4.7 分区	217	8.1.1 MDX 在多维数据集	
6.4.8 透视	220	中的角色	269
6.4.9 翻译	221	8.1.2 使用模板创建	
6.4.10 浏览器	224	第一个 MDX	270
6.5 通过 OLAP 进行数据分析	226	8.1.3 MDX 与 SQL 比较	271
6.5.1 切片	226	8.2 MDX 的构造解析	272
6.5.2 切块	227	8.2.1 MDX 语句的	
6.5.3 钻取	228	基本组成部分	272
6.5.4 旋转	229	8.2.2 成员	273
第7章 数据立方的增强及其应用	231	8.2.3 元组	273
7.1 修改 Cube 结构	231	8.2.4 集合	274
7.1.1 修改度量值	231	8.2.5 区分 MDX 的元素	274

Contents

8.3 MDX 定义语句的使用	275
8.3.1 CREATE SUBCUBE	276
8.3.2 CREATE MEMBER	276
8.4 MDX 操作语句的使用	277
8.4.1 DRILLTHROUGH	277
8.4.2 SELECT	278
8.5 MDX 脚本语句的使用	278
8.5.1 CALCULATE	279
8.5.2 CASE	279
8.5.3 SCOPE	280
8.6 WITH 子句及其使用	281
8.6.1 用 WITH 创建命名集	281
8.6.2 用 WITH 创建计算成员	282
第 9 章 用 SSRS 处理智能报表	283
9.1 报表基本知识	283
9.1.1 报表与商业智能	283
9.1.2 SSRS 的结构	284
9.1.3 SSRS 报表的 3 种状态	285
9.1.4 用 SSRS 做一个 简单的报表	286
9.2 SSRS 的配置与管理	289
9.2.1 报表管理器	290
9.2.2 Reporting Services 配置工具	291
9.2.3 SQL Server 外围 应用配置器	292
9.2.4 SQL Server Management Studio	293
9.2.5 其他配置和管理工具	294
9.3 增强基本报表的功能	294
9.3.1 分组与排序	294
9.3.2 计算	297
9.3.3 参数化报表	297
9.3.4 复杂的参数化报表	299
9.4 报表生成器的使用	302
9.4.1 报表生成器的启动	302
9.4.2 报表模型的创建	302
9.4.3 即席报表的创建	306
第 10 章 基于 SSAS 的商业智能分析	309
10.1 构架商业智能系统的生命周期	309
10.2 商业智能向导	310
10.2.1 商务系统需要哪些智能	311
10.2.2 商业智能向导的 基本使用	311
10.2.3 时间智能	313
10.2.4 账户智能	317
10.3 KPI 分析	319
10.3.1 平衡计分卡 (BSC) 与 KPI	319
10.3.2 KPI 技术在指标 衡量中的作用	320
10.3.3 KPI 分析中的关键问题	321
10.4 80/20 法则的分析	323
10.4.1 80/20 法则	323
10.4.2 计算基本百分比	323
10.4.3 创建百分比比较表	324
10.5 用专业前端展现工具 呈现商业信息	324
10.5.1 前端展现的方法	325
10.5.2 用 Excel 2007 展现数据	326
10.5.3 SharePoint 与 商业智能 portal	336
第 11 章 数据挖掘体系结构与 基本使用方法	339
11.1 数据挖掘基本知识	339
11.1.1 数据挖掘与数据仓库 及商业智能	339
11.1.2 数据挖掘体系结构	341
11.1.3 数据挖掘的工具	341
11.1.4 数据挖掘过程模型	342
11.2 SQL Server 数据挖掘 方案的构成	344
11.2.1 定义问题	344
11.2.2 准备数据	345
11.2.3 浏览数据	345
11.2.4 生成模型	345

11.2.5 浏览和验证模型	346	11.8.2 使用范例	368
11.2.6 部署和更新模型	346	11.9 数据挖掘算法使用小结	370
11.3 一个完整的数据挖掘过程		第 12 章 用数据挖掘技术满足	
——线性回归算法	346	商业分析需求	371
11.3.1 挖掘环境的搭建	346	12.1 一个美丽的爱情故事	371
11.3.2 线性回归的基本原理	348	12.2 商业智能应用中的数据挖掘	372
11.3.3 使用线性回归挖掘数据	349	12.2.1 哪些问题可以使用	
11.4 逻辑回归算法	356	Data mining	372
11.4.1 基本原理	356	12.2.2 用什么挖掘技术	
11.4.2 使用范例	357	解决商业问题	373
11.4.3 验证数据挖掘		12.2.3 挖掘中的非挖掘问题	374
模型的方法	359	12.3 潜在客户分析	375
11.4.4 逻辑回归的参数设置	360	12.3.1 商业需求	375
11.5 决策树算法	361	12.3.2 解决方案	375
11.5.1 基本原理	361	12.3.3 挖掘结果	377
11.5.2 使用范例	362	12.4 购物篮分析	378
11.5.3 决策树的参数设置	363	12.4.1 商业需求	378
11.6 聚类分析算法	364	12.4.2 解决方案	379
11.6.1 基本原理	364	12.4.3 挖掘结果	380
11.6.2 使用范例	365	12.5 数据挖掘的前端展现:	
11.7 Naive Bayes 算法	366	Excel 2007 数据挖掘插件	380
11.7.1 基本原理	366	12.5.1 插件的配置	380
11.7.2 使用范例	366	12.5.2 基本使用方法	381
11.8 关联算法	367	12.6 数据挖掘不是万能的	383
11.8.1 基本原理	367	参考文献	385

第 1 章 发掘数据金矿的工具：

BI 与 DW、OLAP、DM

本章导读：

“工欲善其事，必先利其器”。信息技术在商业领域应用的卓越成效在经过近 20 年的信息化建设已经初步显现。企业通过 MIS（管理信息系统）快速收集和处理商业信息，通过 ERP（企业资源计划系统）准确监控信息流，从而对企业经营的各个方面进行管理。这些系统除了本身的应用外，还积累了大量的数据，如来自业务系统的订单、库存、交易账目、客户和供应商资料，来自企业所处行业和竞争对手的数据，以及来自企业所处的其他外部环境中的各种数据，这是一笔宝贵的财富。信息系统应该具备把这些庞大的数据转化为知识，进而辅助企业经营决策，甚至自动生成商业决策的能力，这就是商业智能。信息系统正在经历着“MIS→ERP→BI”的演变过程。本章将对于商业智能涉及到的相关技术，如数据仓库（DW）、数据挖掘（DM）等进行概要地分析，同时从理论上明确商业智能在企业经营决策过程中的价值发挥原理。

本章先简单地说明企业经营管理活动对商业智能的需求，从而明确实施 BI 的原因，解决 Why 的问题；再阐述商业智能的技术构成，大概了解如何实现商业智能，解决 What 和 How 的问题；最后列举一些实际部署的部署商业智能项目结果，解决 What Effect 的问题。

1.1 企业经营管理活动对商业智能的需求

信息系统在企业管理中的应用不仅随着管理理念的更新而发展，还在很大程度上依赖于企业商业信息的存储量。随着社会节奏的加快，企业信息积累越来越多，而商业知识的需求却得不到满足，于是就提出了商业智能的概念。这一节将对商业智能系统应用的深层原因进行探讨。

1.1.1 企业经营活动面临的挑战

企业经营活动包括采购、生产、销售、商贸磋商、价格比较、经营决策、营销策略、推销促销、公关宣传、售前/售后服务、客户关系和咨询服务等。这些活动的背后实际上有 3 种“流”在支持着整个企业的运转：物流、资金流和信息流。

“物流”是从原材料和零配件到成品整个物资流通过程。它将运输、仓储、加工、整理和配送等整个物资流通过程通过信息有机结合，形成完整的供应和需求链。

“资金流”是资金在企业内部和企业之间的流动过程。它包括融资、应收账款和应付账款的管理等，涉及财务的各个方面。

“信息流”是企业管理的各个环节产生的信息的规律性流动。如业务运作过程产生的单据及其处理过程；经营管理过程中的成本、收入、利润报表及其相互关联；战略决策规划过程中的产品结构及市场定位信息；企业的经营目标和策略信息等。

实际上，每个时期的企业经营活动过程都可以抽象出这“三流”。只不过传统企业管理由于企业规模和经营过程产生的信息量等还不是十分庞大，所以对这“三流”没有进行深入的研究和给予更多的重视。然而，信息时代的企业经营活动在这3个方面却面临着两个很大的挑战。

1. 商务工具的变迁

技术总是不断地影响着商务模式。自从人类社会有商业活动以来，采用新进有效的商业工具往往能在竞争中先发制人。这里不讨论生产工具和交通工具，只着眼于和商业智能关联的信息工具来看现代企业在这方面面临的这一挑战。

每一次科学技术的进步，总是会在很短的时间内在商业领域得到普遍的应用。图 1-1 的上面部分展示了从电报到互联网在不同阶段有代表性的商业工具，伴随着商业工具的变迁，人们处理信息的能力在工具的帮助下变得越来越强大，信息量的增长也越来越快。在此图的下半部分是与商业工具的发展相对应的信息量的增长。可以说，对商业活动影响最大的工具是计算机在商业领域的应用，它标志着商业活动真正进入信息化的时代，这时候的信息处理方式以数据库为主。现阶段企业的很多商业活动的信息化都还属于这个阶段。互联网在商业活动中的应用不仅从某种程度上改变了传统的商业模式，在信息增长和信息处理方式上也发生了前所未有的变化。在这个阶段，伴随着网络的使用，商业信息呈爆炸状增长，企业从以前使用的系统中继承了相当多的数据，如从 ERP 中继承的生产运作相关信息、从 CRM 中继承的客户信息和从 HR 系统中继承得到的员工信息等，这些信息只有共同作用才能发挥其效益，故信息处理方式以数据仓库和数据集市等为主。

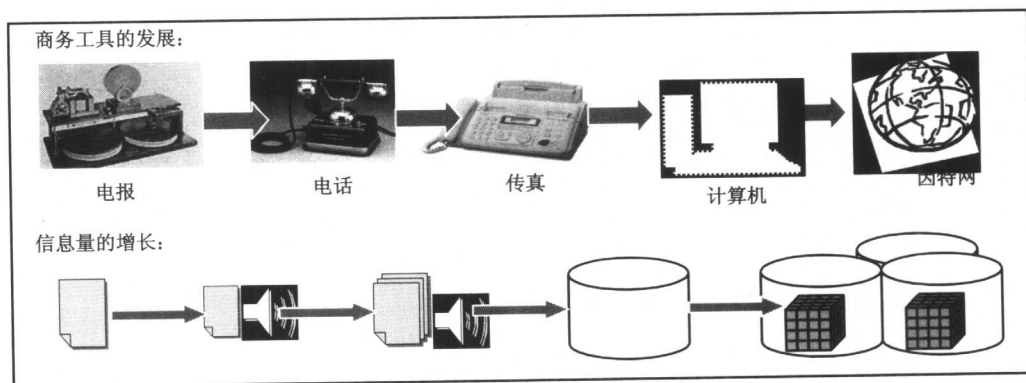


图 1-1 商业工具的发展及其相对应的信息量

正是由于信息工具的变化和信息量的不断增长，人们针对这些不同时期的不同信息类型一直在探索管理好它们的方法，也伴随着产生了一些工具。在使用电报、电话和传真等